UNIVERSAL AND SPECIFIC FEATURES IN INTONATION PERCEPTION

Veronika Makarova (Meikai, university, ETL)

makarova@etl.go.jp

Abstract

This paper reports the results of an experimental phonetic study of intonation contrasts perception by speakers of British English, Japanese and Russian. Six series of re-synthesized two-syllable rise-fall contours with manipulated parameters of the rise in the first and the fall in the second syllable were employed in the experiment. Modifications of pitch height were executed in 2 st steps, and of duration – in 30ms steps. The subjects, who were native speakers of British English, Japanese and Russian, identified the sentence type of presented re-synthesized stimuli. The results of the experiments demonstrate overall similarity of the perception strategies across the three groups of subjects, especially regarding the thresholds of 'declarative' sentence type judgement. Non-declarative judgements are more language-specific. The results can be employed for the teaching of English, Japanese and Russian as foreign languages as well as for speech synthesis and recognition.

1. Introduction

Intonation is responsible for rendering a significant amount of information in spoken discourse. While this feature of intonation is generally believed to be universal, the nature of information contrasts conveyed by intonation as well as the ratio of universal versus language specific in these contrasts are debated. Many researchers agree that intonation can transmit information of two kinds: linguistic and paralinguistic (Ladd, 1996). However, the interpretation of these terms if far from being established. Fujisaki (1997) and Maekawa (1998), for example, believe that paralinguistic information is not inferable from the written

counterpart of an utterance, is deliberately added by the speaker and that it implies continuous rather than categorical contrasts. For other researchers paralinguistic features signify variations in the voice quality and tone which are not controlled by the speaker and are mostly associated with the speaker's individual characteristics, attitudes and emotions (Ladefoged, 1997).

The approach accepted in this paper is based on the assumption that paralinguistic features are not controlled by the speaker, whereas linguistic features are chosen by the speaker from the language inventory to suit the specific information exchange task. Intonation organises spoken text into information units. It ranks information according to its importance, e.g. it highlights the most important words by assigning pitch accents to them. It also indicates by the choice of a particular accent whether each unit contains new or old (given) information (Cahn, 1998). Ensuring turn taking in a conversation, intonation signals the starting and ending points of a message, whether the conveyed information is complete or incomplete and whether any more information is needed. Finally, it demonstrates whether information is given or needs to be retrieved. In other words, it surfaces the syntactic structure and utterance type.

In many world languages the ability of intonation to signal an utterance type has been observed both in speech production and in speech perception. Studies of speech production reveal a link between an utterance (sentence) type and a certain intonation pattern frequently occurring in this type (Cruttenden, 1997). In speech perception experiments, speakers of many different languages, like English (Price, et. al. 1991), Korean (Jang, Song & Lee, 1998), Dutch (Caspers, 1998) have been shown to successfully disambiguate utterance types relying only on prosodic clues. This function of intonation is often believed to be universal (Cruttenden, 1997). However, the exact types of utterance (sentence) type distinctions rendered by intonation contrasts appear to be language-specific.

This paper investigates perception mechanisms which allow the speakers of different languages to identify utterance (sentence) types by intonation means. Some common features in the perception of utterances (sentence) type has been observed across languages. For example, a falling terminal pitch movement tends to be associated with statements, whereas a rising movement or a high pitch with 'yes/no' questions (Cruttenden, 1997). However, some language-specific features have also been found. For example, 'yes/no' questions for the speakers of Japanese and English are associated with a rising pitch movement (Miura & Hara, 1995; Ladd, 1996), whereas for the speakers of Russian – with the rising or a rising-falling movement depending on the presence or absence of post-accented syllables (Svetozarova, 1982).

The experiment reported in this paper is aimed to investigate if changes in the parameters of a rising-falling pitch movement in a short two-syllable contour are associated with the changes of the perceived utterance (sentence) type for the speakers of the three languages: English, Japanese and Russian.

2. Materials and methods

The perception experiment was conducted with six series of re-synthesized speech signals with modified intonation parameters. Modifications and re-synthesis of a real speech signal 'tata' (a rise in the first and a fall in the second vowel) were performed by LPC method (16 coefficients) on a Kay Elemetrics CSL 4300. Obtained stimuli were grouped into 6 series for a perception experiment whereby each series contained one manipulated parameter as follows: 1) height of the ending point of the rise; 2) height of the steep (4.2st) rise; 3) height of the gradual (0.5st) rise; 4) pitch height of the end of the fall and interval of the fall; 5) pitch height of the beginning of the fall and interval of the fall; 6) duration of the rise (short - 55ms, medium – 90ms, long – 135ms) and the fall (60, 120 and 180ms

respectively). All the manipulations of pitch heights were performed in 2 st steps. Figure 1 demonstrates the parameters manipulated in each series.

Forty Japanese and thirty eight Russian subjects, both male and female, performed the forced-choice utterance (sentence) type disambiguation task. The subjects were requested to listen to the six experimental series in succession, and identify each presented stimulus as a statement, a question or an exclamation. The subjects indicated their choice by circling a corresponding sign in the answer sheet. Percent of identification of each stimulus as statement, question or exclamation (also referred to as declarative, interrogative and exclamatory judgement) is shown in Figure 2. Chi-square test was used to determine the significance of the difference in stimuli identification between the three groups of listeners (at $p<0.05$).

3. Results

As can be seen from Figure 2, the results of the listening experiment demonstrate similarities of the listening strategies between the three groups of subjects.

Perception results in Series 1 demonstrate the decrease in declarative, and an increase in interrogative and exclamatory judgements with the increased height of the ending point of the rise for all the three groups of subjects. A threshold of declarative judgement is observed at stimulus 3 (196 to 265 Hz). Subsequent stimuli tend to be perceived as exclamatory by British and Japanese subjects, whereas for Russian subjects they sound either interrogative or exclamatory. These differences between the subjects' performance are statistically significant at stimuli 4 and 5.

Perception results in Series 2 show that raising the starting and ending points of the gradual 0.5 st rise (with a constant interval of the rise) also lead to the decrease in declarative and an increase in non-declarative judgement by all the listeners. However, percent of

exclamatory perception of the stimuli is higher for the Japanese and British than for Russian subjects.

Perception results in Series 3 (raising the starting and ending points) are similar to the results in Series 2. The thresholds of declarative judgement in Series 2 (Stimulus 3: 175 to 223 Hz) and in Series 3 (Stimulus 4: 216 to 223 Hz) have the same height of the ending point (223 Hz). This suggests that the height of the ending point of the rise is a more salient factor for the listeners' perception than the height of the starting point and/or the interval of the rise.

Listeners' responses in Series 4 (raising the height of the fall end) demonstrate a small effect of this parameter on listeners' perception as compared to the parameters of the rise manipulated in series 1-3. Raised height of the fall end leads to a very gradual decrease in the British and Japanese listeners' declarative judgement, and has no effect on the perception of Russian subjects (this difference is significant at Stimulus 1: 217 to 72 Hz).

In contrast to the height of the ending point of the rise, the height of the starting point of the rise manipulated in Series 5 appears to have more effect on listeners' perception whereby the increase in pitch height decreases declarative and increases exclamatory judgements for all the three groups of listeners.

Results in Series 6 (manipulated duration) show that this factor has the least effect on listeners' perception of utterance (sentence) type. All the stimuli in this series tend to be perceived as statements. Japanese and Russian subjects give the highest declarative judgement to stimulus 4 (short+medium) whereas British subjects prefer stimuli 1,6,8 (short+short, long+medium and medium+long).

4. Conclusion

The results of the experiment suggest that the speakers of unrelated or distantly related languages with different types of prosodic systems and different employment of intonation

contrasts for utterance/sentence type disambiguation can rely on the same clues while performing an utterance/sentence type disambiguation task and show similar perception strategies. The parameters of the rising part of the contour, especially the height of the ending point of the rise were an important clue for the speakers of English, Japanese and Russian. Parameters of the fall were of lesser, and of duration – of little salience for the perception of utterance/sentence type. These results support the idea of cross-linguistic universals in intonation perception.

On the other hand, some observed differences across the subject groups, e.g., in the perception of duration and manipulated height of the fall end can be explained by the language-specific features in intonation functions, more specifically by the expectations of the listeners formed by their native language intonation systems. Larger percent of interrogative judgements given by Russian subjects as compared to British and Japanese subjects is also explained by the differences between the language systems. Russian language system frequently exposes Russian subjects to short questions with a rise-fall contour and no lexico-grammatical markers of sentence types, whereas in British English rise-fall questions are commonly used, but are on the periphery of the linguistic consiousness of the speakers. In Japanese rise-fall questions with a rise-fall contour are found only very rarely in echoes.

Series 1-3 show the existence of declarative judgment thresholds whereby a gradual decrease in declarative judgement is followed by a sharp drop at the threshold. These results suggest that the quantum theory of speech perception developed by Stevens (1972) for segmental phonetics can also hold true for the perception of intonation contrasts.

Acknowledgements

This paper became possible due to Speech Signal Processing project conducted at the Electrotechnical Laboratory. I would like to express my special gratitude to Dr. Kazuyo

Tanaka for his advise and support.

<REFERENCES>

Cahn, J. E. (1998). A computational memory and processing model for prosody. In Mannell, R. and Robert-Ribes, J. (Eds) *ICSLP' 1998. Proceedings.* V. 3, pp.579-582.

Caspers, J. (1998). Who's next? The melodic marking of question versus continuation in Dutch. *Language and Speech*, Vol. 41, N 3-4, pp. 375-398.

Cruttenden, A. (1997). *Intonation.* 2d ed. Cambridge: Cambridge University Press.

Fujisaki, H. (1997). Prosody, models and spontaneous speech. In Sagisaka et al. (Ed). *Computing Prosody*, Springer.

Horiuchi, Y., Ichikawa, A. Prosodic structure in Japanese spontaneous speech. In Mannell, R. and Robert-Ribes, J. (Eds) *ICSLP' 1998. Proceedings.* V. 3, pp.591-594.

Jang, T.Y., Song, M., Lee, K. (1998). Disambiguation of Korean utterances using automatic intonation recognition. In Mannell, R. and Robert-Ribes, J. (Eds) *ICSLP' 1998. Proceedings.* V. 3, pp.603-606.

Koike, K., Suzuki, H., Saito, H. (1998). Prosodic parameters in Emotional speech. In: Mannell, R. and Robert-Ribes, J. (Eds) *ICSLP' 1998. Proceedings.* V. 3, pp.679-682.

Ladd, D.R. (1996). *Intonational Phonology*, Cambridge: CUP.

Ladefoged, P. (1997). Linguistic phonetic descriptions. In: W. J. Hardcastle and J. Laver (Eds). *The handbook of phonetic sciences*. Oxford: Blackwell, pp. 589-618.

Maekawa K. (1998). Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. In Mannell, R. and Robert-Ribes, J. (Eds) *ICSLP' 1998. Proceedings.* V. 3, pp.635-638.

Miura, I. and Hara, N. (1995). Production and perception of rhetorical questions in Osaka Japanese. *Journal of Phonetics*, V. 23, pp. 291-303.

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustic Society of America*, V. 90 (6): 2956-2970.

Stevens, K. N. (1972). The quantal nature of speech: Evidence from articulatory-acoustic data. In: David, E. E. and Denes, P. B., (Ed.), *Human Communication: A Unified View*, New York: McGraw-Hill, , pp. 51-66.

Svetozarova, N.D. (1982). *Intonatsionnaya sistema russkogo yazyka*, Leningrad: Leningrad University.

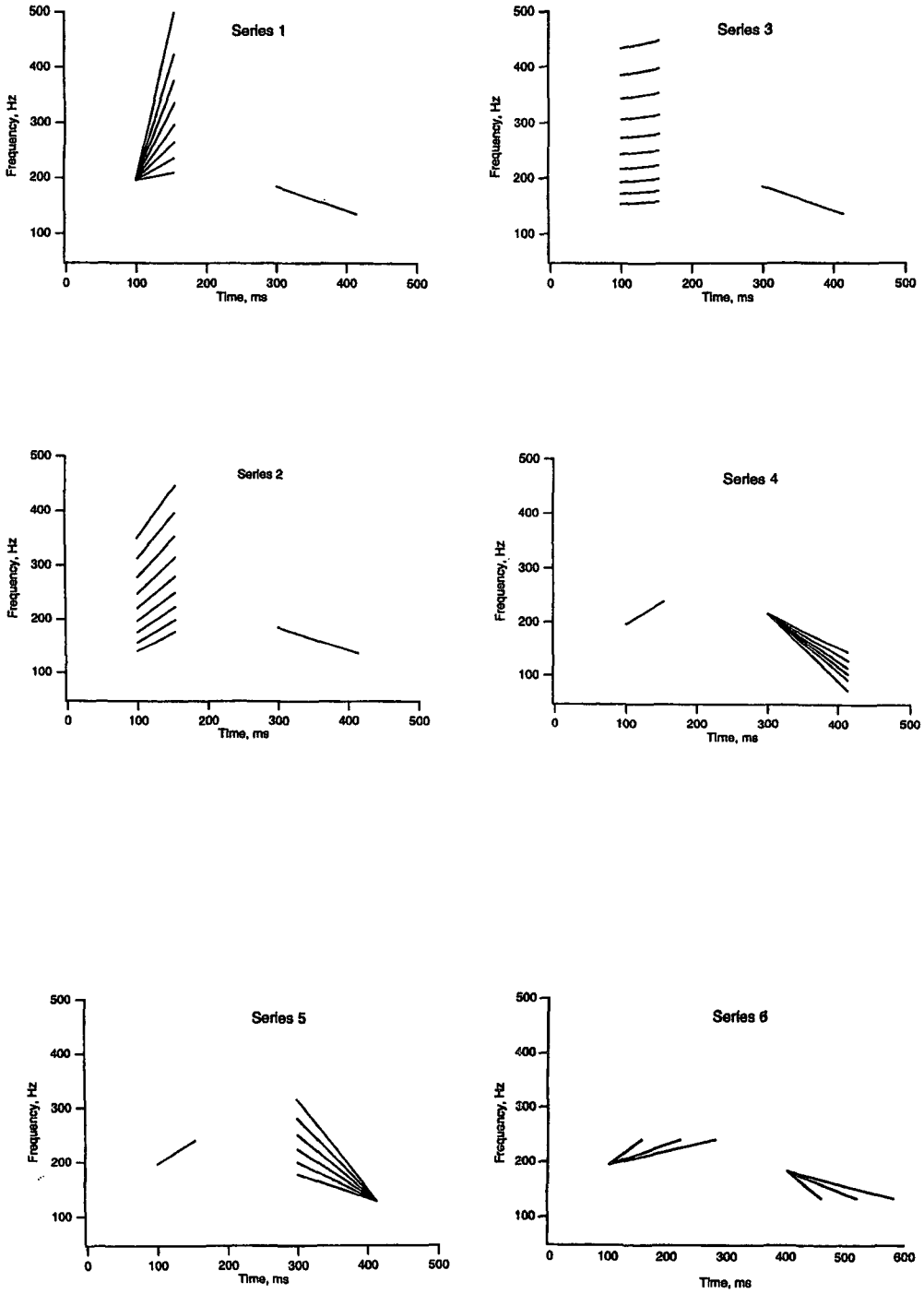

Figure 1.

Manipulated parameters in Series 1-6

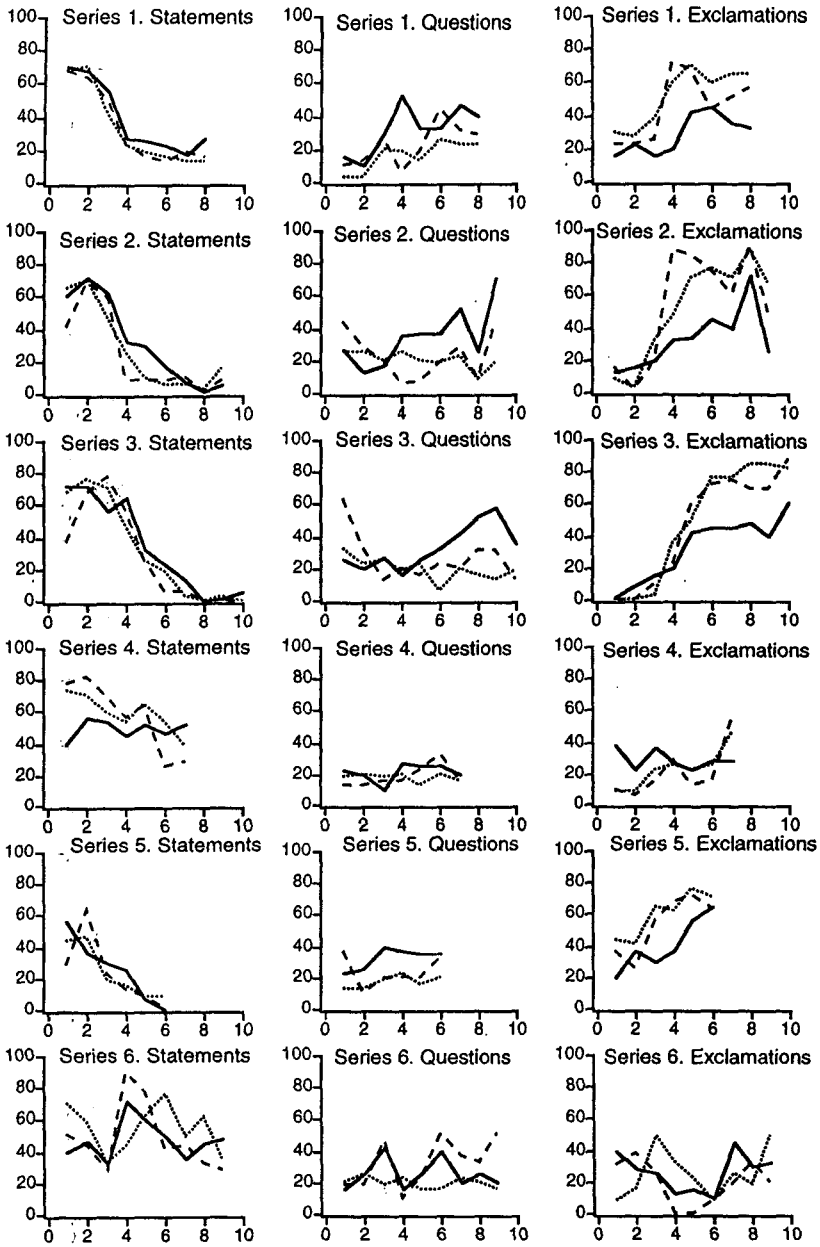

Figure 2: Listeners' perception of the stimuli

Vertical axis indicate percent of stimuli identification as belonging to one of the three sentence types; horizontal axis indicate numbers of stimuli (from the lowest to the highest). Responses of Russian subjects are shown by solid lines, of Japanese subjects – by dashed lines, and of British subjects – by dotted lines.